PART B: REGRESSION (50 marks)

This part of the assignment is to be completed individually.

Background

This is a dataset to predict the housing price in US based on various factors such as city, house area, No. of bedrooms and toilets, renovation status, etc.

Dataset

You are going to use the dataset: **housing_dataset.csv**.

Tasks

1. Write the code to solve the prediction task. You should use scikit-learn for the machine learning models (no 3rd party libraries).

2. **In the Jupyter notebook**, write your report detailing your implementation, your experiments and analysis (along with your python code and comments). In particular, we'd like to know:

   - How is your prediction task defined? And what is the meaning of the output variable?
   - How do you represent your data as features?
   - Did you process the features in any way?
   - How did you select which learning algorithms to use?
   - Did you try to tune the hyperparameters of the learning algorithm, and in that case how?
   - How do you evaluate the quality of your system?
   - How well does your system compare to a dummy baseline?
   - Can you say anything about the errors that the system makes?
   - Is it possible to say something about which features the model considers important?

3. Create a set of slides with the highlights of your Jupyter notebook report. Explain the entire machine learning process you went through, data exploration, data cleaning, feature engineering, model building and evaluation, and model improvement. Write your conclusions.